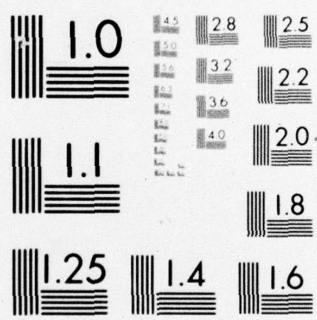
AD-A041 686 ILLINOIS UNIV AT URBANA-CHAMPAIGN COORDINATED SCIENCE LAB F/G 9/4
VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING SUB--ETC(U)
MAY 77 K M MACKENTHUN DAAB07-72-C-0259
UNCLASSIFIED RT-767 NL

| OF |
AD
A041686

END
DATE
FILMED
8-77



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

ADA041686

REPORT R-767 MAY, 1977 UILU-ENG 77-2214

CSL COORDINATED SCIENCE LABORATORY

VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING SUBJECT TO A FIDELITY CONSTRAINT

KENNETH MARSH MACKENTHUN, JR.

UNIVERSITY OF ILLINOIS - URBANA, ILLINOIS

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

12

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING SUBJECT TO A FIDELITY CONSTRAINT

5. TYPE OF REPORT & PERIOD COVERED

9 Technical Report,

6. PERFORMING ORG. REPORT NUMBER

R-767, UILU-ENG-77-2214 14

7. AUTHOR(s)

10 Kenneth Marsh Mackenthun, Jr.

8. CONTRACT OR GRANT NUMBER(s)

15 DAAB 07-72-C-0259, NSF-ENG-75-20864

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

12 59 p.

11. CONTROLLING OFFICE NAME AND ADDRESS

Joint Services Electronics Program

12. REPORT DATE

11 May 1977

13. NUMBER OF PAGES

50

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Rate Distortion Theory
Information Theory
Universal Source Coding

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The goal of universal source coding is to find codes which are efficient for sources with unknown or incompletely specified probability distributions; this is equivalent to finding codes which are efficient for a collection of sources. Previous work in universal variable-rate source coding subject to a fidelity constraint has assumed a prior probability measure on a collection of stationary ergodic sources. It is then shown that there is a single code which gives an average distortion D at an optimum rate, $R_\theta(D)$, for most

RSUB THETA Cont

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 097 700 mt

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

THETA R SUB THETA

20. ABSTRACT (continued)

sources θ in the collection. This code is necessarily variable-rate since $R_\theta(D)$ is not constant over the collection of sources.

In this dissertation, new results are presented on coding a collection of stationary ergodic sources when no prior probability measure is assumed. In particular, it is shown that there exist variable-rate universal codes whose rate-distortion performance converges to the optimum attainable performance for each source in the collection. Two types of universal codes are considered in which the above convergence is pointwise (weakly-universal) or uniform (strongly-universal) over the collection of sources. A major result of the work is that weakly-universal codes exist for a collection of stationary ergodic sources with source and reproducing alphabets separable metric spaces and the distortion measure a nonnegative nondecreasing function of the metric. It is assumed that there is a letter in the reproducing alphabet giving finite average distortion when used with each source in the collection and that $\sup\{R_\theta(D)\}$ over the collection of sources of finite. Strongly-universal codes are shown to exist for collections of finite alphabet memoryless sources and certain types of Markov sources.

END

SEE ABOVE

UILU-ENG 77-2214

VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING SUBJECT TO A FIDELITY CONSTRAINT

by

Kenneth Marsh Mackenthun, Jr.

This work was supported in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract DAAB-07-72-C-0259 and in part by the National Science Foundation under Grant NSF ENG 75-20864.

VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING SUBJECT TO A FIDELITY CONSTRAINT

BY

KENNETH MARSH MACKENTHUN, JR.

B.A., Case Western Reserve University, 1972
B.S., George Washington University, 1974
M.S., University of Illinois, 1976

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1977

Thesis Adviser: Professor Michael B. Pursley

Urbana, Illinois

VARIABLE-RATE, WEAKLY- AND STRONGLY-UNIVERSAL SOURCE CODING

SUBJECT TO A FIDELITY CONSTRAINT

Kenneth Marsh Mackenthun, Jr., Ph.D.
Coordinated Science Laboratory and
Department of Electrical Engineering
University of Illinois at Urbana-Champaign, 1977

Abstract

The goal of universal source coding is to find codes which are efficient for sources with unknown or incompletely specified probability distributions; this is equivalent to finding codes which are efficient for a collection of sources. Previous work in universal variable-rate source coding subject to a fidelity constraint has assumed a prior probability measure on a collection of stationary ergodic sources. It is then shown that there is a single code which gives an average distortion D at an optimum rate, $R_\theta(D)$, for most sources θ in the collection. This code is necessarily variable-rate since $R_\theta(D)$ is not constant over the collection of sources.

In this dissertation, new results are presented on coding a collection of stationary ergodic sources when no prior probability measure is assumed. In particular, it is shown that there exist variable-rate universal codes whose rate-distortion performance converges to the optimum attainable performance for each source in the collection. Two types of universal codes are considered in which the above convergence is pointwise (weakly-universal) or uniform (strongly-universal) over the collection of sources. A major result of the work is that weakly-universal codes exist for a collection of stationary ergodic sources with source and reproducing alphabets separable metric spaces and the distortion measure a nonnegative

nondecreasing function of the metric. It is assumed that there is a letter in the reproducing alphabet giving finite average distortion when used with each source in the collection and that $\sup\{R_\theta(D)\}$ over the collection of sources is finite. Strongly-universal codes are shown to exist for collections of finite alphabet memoryless sources and certain types of Markov sources.

Acknowledgements

I would like to thank my thesis adviser, Dr. M. B. Pursley, for his patience, support, encouragement, and many helpful comments during the research period. I would like to thank Drs. A. H. Haddad, F. P. Preparata, and R. B. Ash for serving on the doctoral examination committee.

Many thanks go to Phyllis Young for her endurance and expertise during the typing of this manuscript.

Table of Contents

Chapter 1. Introduction

Shannon's classical papers in information theory [1,2] deal with the coding of a source with known probability distribution. However, in practice, the probability distribution of the source is rarely (if ever) known precisely, and yet it is desired to find a code which performs asymptotically as well as a code designed with full knowledge of the source probability distribution. We will refer to this as the universal coding problem.

We assume in the treatment of this problem that the reproducing alphabet $\hat{A}_0$ used to encode the source is known and that the source alphabet is a subset of some known set A_0. If nothing is known about the source probability distribution, then the universal coding problem has no solution since the unknown source could be ergodic, stationary, or non-stationary and the universal code would be required to work well in each of these cases. Therefore, it is necessary to give some information about the probability distribution of the unknown source; in our discussion and the great majority of the literature, this is done by specifying some regularity property of the source, i.e., it is stated that the unknown source is memoryless, Markov, stationary ergodic, stationary, etc. With this and sometimes additional information, interesting results can be obtained on the universal coding problem.

It is now possible to give an equivalent formulation of the universal coding problem. The unknown source is assumed to be a member of the class of all sources having the same regularity property and other properties assumed for the unknown source; sources in the class are indexed

by a parameter θ taking values in a set Λ. The universal coding problem is then equivalent to the problem of finding a single code which works well for each member of the class of sources with index set Λ. Henceforth, we discard the notion of an unknown source and instead refer to this equivalent formulation when discussing the universal coding problem. In our discussion, we seek only universal codes which encode blocks of source outputs independently, that is, the encoding of a particular block output does not depend on the source outputs during any previous blocks.

Universal codes can be roughly divided into whether the code reproduces the source perfectly (noiseless coding) or whether distortion is allowed (coding with a fidelity criterion). We will be concerned with only the latter type of code in this work. Universal noiseless coding has been discussed by Fitingof [3], Ziv [4], and Davisson [5].

There are two distinct philosophies in universal coding with a fidelity criterion. The most prevalent approach and historically the first has been to use fixed length block codes of some fixed rate R to achieve an optimum average distortion, $D_\theta(R)$, for most $\theta \in \Lambda$. Since these codes guarantee a transmission rate R for each $\theta \in \Lambda$, they can be used to transmit over a fixed rate channel if the user is willing to tolerate the large average distortions that may result for some $\theta \in \Lambda$.

There are several different categories of fixed rate universal codes. If W is any weighting measure assigned to the set Λ and C is a fixed length block code of rate R whose average distortion for the source with index θ is $\rho(C,\theta)$ then we say C is a <u>weighted universal</u> fixed rate

code if

$$\int_\Lambda \rho(C,\theta)W(d\theta) \approx \int_\Lambda D_\theta(R)W(d\theta).$$

Such codes were first discussed by Gray and Davisson [6], where it was shown that weighted universal codes exist for A_0 a separable metric space, $\hat{A}_0$ a totally bounded metric space, the distortion measure ρ a metric on $A_0 \cup \hat{A}_0$, and Λ an index set for the class of all stationary ergodic sources on A_0. If μ is a stationary source with source alphabet A_0, then the ergodic decomposition theorem [7] allows us to identify W with μ and Λ with A_0 and thereby show that the weighted average distortion of the universal code is the optimum distortion attainable by any code of the same rate applied to the stationary source [6]. Another result which follows quickly from the existence of weighted universal codes is that for any $\varepsilon > 0$, there is a set $\Lambda_0 \subset \Lambda$ with $W(\Lambda_0) \geq 1 - \varepsilon$ and a single fixed length block code C of rate R for which

$$\rho(C,\theta) \approx D_\theta(R)$$

for all $\theta \in \Lambda_0$. Such a code is universal in the sense that it gives an optimum average distortion for "most" $\theta \in \Lambda$.

Two further types of universal codes, <u>weakly minimax</u> and <u>strongly minimax</u>, dispense with the assumption of a prior weighting measure on the class Λ [8]. Weakly minimax codes exist for Λ if there is a sequence $\{C_n\}$ of fixed length block codes of fixed rate R such that $\rho(C_n,\theta) \to D_\theta(R)$ for each $\theta \in \Lambda$. Strongly minimax universal codes exist for Λ if there is a

single fixed length block code C of rate R such that $\rho(C,\theta) \approx D_\theta(R)$ for all $\theta \in \Lambda$. Ziv first showed the existence of weakly minimax universal codes for general sources and alphabets [9]. His results were generalized further by Neuhoff, Gray, and Davisson [8] who showed that weakly minimax universal codes exist for the class of all stationary ergodic sources with A_0 and $\hat{A}_0$ metric spaces such that either A_0 or $\hat{A}_0$ is separable and the distortion measure ρ a metric on $A_0 \cup \hat{A}_0$. In [8], strongly minimax universal codes were shown to exist for a class of sources totally bounded under the $\bar{\rho}$ distance measure [10] if A_0 is finite or a totally bounded metric space. An example of a class of sources totally bounded under the $\bar{\rho}$ distance is a class of memoryless sources on a metric space alphabet and a class of finite alphabet n-th order Markov sources having transition probabilities bounded away from zero. In [11], Neuhoff and Shields discuss strongly minimax coding for Markov sources. They show that strongly minimax codes exist for the class of all binary 1st-order Markov sources, and that for 1st-order Markov sources with source alphabets of K symbols, $K > 2$, strongly minimax codes of rate R exist if and only if $R \geq \log(K-1)$.

The basic motivation behind the second approach to universal coding is to find universal codes which guarantee average distortion $\approx D$ for all $\theta \in \Lambda$. An efficient universal code which guarantees average distortion $\approx D$ for each $\theta \in \Lambda$ should have average rate $\approx R_\theta(D)$ for each $\theta \in \Lambda$. It is clear that to meet these requirements when $R_\theta(D)$ is not a trivial function of θ, a variable rate code must be used. To accomplish this, a binary interface

is inserted between the source output block and the block of reproduction letters. The binary interface, actually a binary variable length prefix condition code book, is then designed so that for each $\theta \in \Lambda$ the encoding and decoding mapping preserves the distortion properties of a good block code specifically designed for that θ and so that the average transmission rate of the binary code is close to $R_\theta(D)$. Pursley and Davisson [12] first took this approach to universal coding and obtained results in the variable rate sense which roughly correspond to weighted universal fixed rate results. In the variable rate terminology, we say that the class of sources is W-almost universally encodable if for each $\epsilon > 0$ there exists a binary variable rate code $\underline{B}$ and a set $\Lambda_0 \subset \Lambda$ with $W(\Lambda_0) \geq 1 - \epsilon$ such that for each $\theta \in \Lambda_0$,

$$\bar{\rho}(\underline{B},\theta) \leq D + \epsilon$$

and

$$\bar{r}(\underline{B},\theta) \leq (\ln 2)^{-1}[R_\theta(D) + \epsilon].$$

Here $\bar{\rho}(\underline{B},\theta)$ and $\bar{r}(\underline{B},\theta)$ are the average distortion and average rate of $\underline{B}$ for $\theta \in \Lambda$. In [12], Pursley and Davisson showed that the class of all stationary ergodic sources with A_0 a separable metric space, $\hat{A}_0$ a totally bounded metric space and ρ a metric on $A_0 \cup \hat{A}_0$ is W-almost universally encodable. In [13], this result was extended to allow $\hat{A}_0$ to be either a Hilbert space or a metric space for which every closed bounded subset is compact. Subsequently, Kieffer [14] proved the result for $\hat{A}_0$ a separable metric space and ρ a nonnegative nondecreasing function of the underlying metric.

One potential use for variable rate universal codes would be in the transmission of satellite weather photographs. The weather images are often stored in digital form in a computer memory or on magnetic tape. In converting the original image to digital form, distortion is necessarily introduced. An efficient universal variable rate code would guarantee a fixed average distortion in the digital reproduction process and allow the video information to be stored on the minimum amount of magnetic tape or computer memory.

In this dissertation, we seek to obtain variable rate universal coding results for general sources and distortion measures without the use of a prior weighting measure on Λ. We define two types of universal variable rate codes which are analogous to fixed rate weakly and strongly minimax universal codes. New results on these types of universal codes are presented in Chapters 3 and 4, and in Appendices A and B, the proofs of results in Chapters 3 and 4, respectively, are presented. Chapter 2 contains necessary mathematical details for the subsequent discussion.

Chapter 2. Mathematical Preliminaries

We let $\mathcal{A}_0$ denote a σ-field of subsets of the source alphabet A_0. If there are no other restrictions on A_0, such as topological restrictions (e.g., metrizability or separability) or restrictions on the cardinality, then we will refer to A_0 as an abstract alphabet. For $n > m$ where n and m are integers or $\pm\infty$, define $A_m^n = \overset{n-1}{\underset{t=m}{\times}} A_t$ and $\mathcal{A}_m^n = \overset{n-1}{\underset{t=m}{\times}} \mathcal{A}_t$ where $A_t = A_0$ and $\mathcal{A}_t = \mathcal{A}_0$ for $m \le t \le n-1$. We write A for $A_{-\infty}^{\infty}$ and $\mathcal{A}$ for $\mathcal{A}_{-\infty}^{\infty}$. We let $\mathcal{F}_{-\infty}^{0}$ denote the σ-field of all sets of the form $G \times A_1^{\infty}$ where $G \in \mathcal{A}_{-\infty}^{0}$ and for each positive integer n we let $\mathcal{F}_n^{\infty}$ denote the σ-field of all sets of the form $A_{-\infty}^{n-1} \times G$ where $G \in \mathcal{A}_n^{\infty}$. For $\omega = \ldots,\omega_{-1},\omega_0,\omega_1,\ldots \in A$ and integers t, define $X_t\colon A \to A_0$ by $X_t(\omega) = \omega_t$. Let T be the shift transformation on A. For positive integers k, define $\underline{X}^k\colon A \to A_0^k$ by $\underline{X}^k = (X_0,X_1,\ldots,X_{k-1})$. Denote elements of A_0^n by $\underline{x}^n = (x_0,x_1,\ldots,x_{n-1})$. For $n = jk$ where j and k are integers, let $(\underline{x}^n)_{i,k} = (x_{ik},x_{ik+1},\ldots,x_{ik+k-1})$ for $0 \le i \le j-1$; i.e., $(\underline{x}^n)_{i,k}$ is the (i+1)-st k vector of $\underline{x}^n$.

For reproduction space $(\hat{A}_0,\hat{\mathcal{A}}_0)$, let $\rho\colon A_0 \times \hat{A}_0 \to [0,\infty)$ be an $\mathcal{A}_0 \times \hat{\mathcal{A}}_0$-measurable function, and define $\rho_n\colon A_0^n \times \hat{A}_0^n \to [0,\infty)$ by

$\rho_n(\underline{x}^n,\underline{y}^n) = n^{-1} \sum_{i=0}^{n-1} \rho(x_i,y_i)$. We will often abbreviate $\rho_n(\underline{x}^n,\underline{y}^n)$ by $\rho(\underline{x}^n,\underline{y}^n)$.

Let $R_\theta(\cdot)$ be the rate distortion function (in nats) for the stationary source $(A,\mathcal{A},\mu_\theta)$, distortion measure ρ, and reproduction alphabet $\hat{A}_0$. Let $\{(A,\mathcal{A},\mu_\theta) \mid \theta \in \Lambda\}$ denote an arbitrary collection of such sources.

If C is any finite set, let $|C|$ be the cardinality of C. Let $[C]^k$ be the k-fold product set $C \times C \times \ldots C$. If $C \subset \hat{A}_0^n$, we call C a <u>block code of length n</u>. For $\underline{x}^n \in A_0^n$, let

$$\rho_n(\underline{x}^n|C) = \min\{\rho_n(\underline{x}^n,\underline{y}^n)|\underline{y}^n \in C\}.$$

Following [12], a <u>binary variable-rate code on blocks of length n</u>, $\underline{B}_n$, is a triple (B,U,V), consisting of codebook B, a finite set of binary sequences satisfying the prefix condition; an $\mathcal{A}_0^n$-measurable encoding mapping $U: A_0^n \to B$; and a decoding mapping $V: B \to \hat{A}_0^n$. Let $\ell: B \to \{1,2,\ldots\}$ be the function which assigns to each codeword in B its length in bits. The average distortion for code $\underline{B}_n$ applied to the θ-th source is

$$\bar{\rho}_n(\underline{B}_n,\theta) = E_\theta \rho_n(\underline{X}^n, V[U(\underline{X}^n)])$$

and average rate (in bits per source symbol) is

$$\bar{r}_n(\underline{B}_n,\theta) = E_\theta n^{-1} \ell[U(\underline{X}^n)]$$

where E_θ is the expectation with respect to source $(A,\mathcal{A},\mu_\theta)$; that is,

$$E_\theta f = \int_A f\, d\mu_\theta$$

for any $\mathcal{A}$-measurable, real-valued function f. When the code $\underline{B}_n$ is clear from context we will often denote $\bar{\rho}_n(\underline{B}_n,\theta)$ and $\bar{r}_n(\underline{B}_n,\theta)$ by $\bar{\rho}_n(\theta)$ and $\bar{r}_n(\theta)$.

As discussed in [12], any block code C of length n has a natural representation as a binary variable-rate code, which we denote $\underline{B}(C)$. In this representation, if $\underline{B}(C) = (B,U,V)$, then

$$\rho_n(\underline{x}^n, V[U(\underline{x}^n)]) = \rho_n(\underline{x}^n|C) \tag{1}$$

and

$$\ell[U(\underline{x}^n)] = \lceil \log_2|C| \rceil \tag{2}$$

for all $\underline{x}^n \in A_0^n$ where $\lceil \log_2|C| \rceil$ is the smallest integer greater than or equal $\log_2|C|$.

The following statement of the source coding theorem will be used frequently in the remainder of the dissertation. This result is proven in [15] for finite source alphabets, but the proof extends to abstract source alphabets using techniques of [16, Ch. 7], as stated in [15].

Theorem 1: Let $(A,\mathcal{A},\mu_\theta)$ be a stationary ergodic source and assume there is a letter $b \in \hat{A}_0$ for which $E_\theta\, \rho(X_0,b) < \infty$. Then for any $\delta > 0$ and any $D \geq 0$ for which $R_\theta(D) < \infty$, there is an integer N_δ such that for $n \geq N_\delta$, there is a block code C of length n, such that $\underline{b}^n = (b,b,\ldots,b) \in C$,

$$\mu_\theta\{\rho(\underline{X}^n|C) > D + \delta\} < \delta,$$

and

$$n^{-1}\lceil \log_2|C| \rceil < (\ln 2)^{-1}[R_\theta(D) + \delta].$$

Chapter 3. Weakly Universal Variable Rate Codes

In this chapter we consider weakly universal source coding which is defined as follows.

Definition 1. A family $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ of stationary ergodic sources is weakly universally encodable at distortion level D with reproduction alphabet $\hat{A}_0$ if for each $\epsilon > 0$ there exists a sequence $\{\underline{B}_n\}$ of binary variable-rate codes (called an ϵ-weakly universal sequence) such that for each $\theta \in \Lambda$ there exists an integer $N(\theta)$ for which $n \geq N(\theta)$ implies

$$\bar{\rho}_n(\underline{B}_n,\theta) \leq D + \epsilon$$

and

$$\bar{r}_n(\underline{B}_n,\theta) \leq (\ln 2)^{-1}[R_\theta(D) + \epsilon].$$

Loosely speaking, an ϵ-weakly universal sequence is the variable-rate counterpart of the weakly-minimax universal codes of [10]. For each $\theta \in \Lambda$, weakly-minimax universal codes have a fixed rate R and an average distortion which approaches the optimum theoretically attainable, $D_\theta(R)$, whereas an ϵ-weakly universal sequence has an average distortion which approaches D and an average rate which approaches the minimum theoretically attainable rate, $R_\theta(D)$.

Our key result is the following theorem.

Theorem 2. Suppose A_0 is an abstract alphabet and $\hat{A}_0$ is countable. Suppose also that the following two conditions hold.

$$\exists\, b \in \hat{A}_0 \ni E_\theta\, \rho(X_0,b) < \infty, \quad \forall\, \theta \in \Lambda \tag{3}$$

$$\sup\{R_\theta(D) | \theta \in \Lambda\} < \infty \tag{4}$$

Then the class $\{A,\mathcal{A},\mu_\theta | \theta \in \Lambda\}$ is weakly universal encodable at distortion level D with reproduction alphabet $\hat{A}_0$.

Remarks. Notice that if A_0 is a metric space, (3) is equivalent to

$$\forall\ \theta \in \Lambda\ \exists\ b_\theta \in A_0 \ni E_\theta\, \rho(X_0,b_\theta) < \infty. \tag{5}$$

For applications, the two theorems which follow are considerably more important than Theorem 2; both theorems use the following condition.

Condition (a). There is a countable set $A_0' \subset \hat{A}_0$ such that for each $\epsilon > 0$, each $\theta \in \Lambda$, and each finite set $C \subset \hat{A}_0^n$, there exists $C' \subset (A_0')^n$ such that $|C| = |C'|$ and $E_\theta \rho(\underline{X}^n|C') \le E_\theta \rho(\underline{X}^n|C) + \epsilon$. In [14], Kieffer shows that the following two important examples satisfy Condition (a).

Example 1. $\hat{A}_0$ is a separable metric space, ρ is bounded, and $\rho(x,\cdot)$ is continuous for each $x \in A_0$.

Example 2. There exists a nondecreasing continuous function $f:[0,\infty) \to [0,\infty)$ such that for each $a > 0$, $\lim\limits_{x\to\infty} [f(x+a)/f(x)] = 1$ and there exists a metric d on $A_0 \cup \hat{A}_0$ such that $\hat{A}_0$ is separable and $\rho(x,y) = f[d(x,y)]$ for each $x \in A_0$ and $y \in \hat{A}_0$.

Theorem 3. Suppose A_0 and $\hat{A}_0$ are abstract alphabets and $\{(A,\mathcal{A},\mu_\theta | \theta \in \Lambda\}$ is a class of stationary ergodic sources such that (3), (4), and Condition (a) are all satisfied. Then the class $\{(A,\mathcal{A},\mu_\theta)| \theta \in \Lambda\}$ is weakly universally encodable at distortion level D with reproduction alphabet $\hat{A}_0$.

As another consequence of Theorem 2, we obtain the following result, first proved by Kieffer [14], which is the most general almost-universal coding theorem. It includes all of the theorems of [12] and [13] as special cases.

Theorem 4. Suppose A_0 is an abstract alphabet, $(\Lambda,\mathcal{L},W)$ is a probability space for which the mapping $\theta \to \mu_\theta(E)$ is $\mathcal{L}$-measurable for each $E \in \mathcal{A}$, and Condition (a) is satisfied. Suppose also there is a $\Lambda' \in \mathcal{L}$ such that $W(\Lambda') = 1$ and such that the following two conditions are satisfied:

$$\exists\, b \in \hat{A}_0 \ni E_\theta\, \rho(X_0,b) < \infty, \quad \forall\, \theta \in \Lambda'$$

and

$$R_\theta(D) < \infty, \quad \forall\, \theta \in \Lambda'.$$

Then the class $\{(A,\mathcal{A},\mu_\theta)\,|\,\theta \in \Lambda\}$ is W-almost universally encodable at distortion level D with reproduction alphabet $\hat{A}_0$.

It is desirable to weaken (4) if possible since there are interesting classes of sources which do not satisfy this. However, we note that (4) arises naturally in many important situations.

Example 3. $\hat{A}_0$ finite and $R_\theta(D) < \infty$ for all $\theta \in \Lambda$.

Example 4. $\hat{A}_0$ totally bounded under metric ρ and $R_\theta(D) < \infty$ for all $\theta \in \Lambda$. In this case, for each $\epsilon > 0$ we can find a finite approximating set $A^\epsilon \subset \hat{A}_0$ for which $R_\theta(D + \epsilon, A^\epsilon) \le R_\theta(D)$ for all $\theta \in \Lambda$, where $R_\theta(D + \epsilon, A^\epsilon)$ is the rate-distortion function for reproducing alphabet A^ϵ [12], [13]. Since $R_\theta(D) < \infty$ for all $\theta \in \Lambda$, it follows that $R_\theta(D + \epsilon, A^\epsilon)$ is given by

an information-theoretic minimization and therefore $R_\theta(D + \epsilon, A^\epsilon) \le \ln |A^\epsilon|$ for all $\theta \in \Lambda$. Thus (4) is satisfied and applying Theorem 3, we can find an ϵ-weakly universal sequence for Λ and A^ϵ which is 2ϵ-weakly universal for Λ and $\hat{A}_0$.

Example 5. Compact class of sources and $R_\theta(D) < \infty$ for all $\theta \in \Lambda$. If the class of sources is compact [17], then for any $\epsilon > 0$, there is a totally bounded set $A^\epsilon \subset \hat{A}_0$ and a mapping $\tau: A^\epsilon \to \hat{A}_0$ such that $E_\theta\rho(X_0, \tau(X_0)) < \epsilon$ for all $\theta \in \Lambda$. Under this condition, it is easy to see that $R_\theta(D + \epsilon, A^\epsilon) \le R_\theta(D)$ for all $\theta \in \Lambda$. Since A^ϵ is totally bounded under the metric ρ and $R_\theta(D) < \infty$ for all $\theta \in \Lambda$, we can apply the result of Example 4 and obtain an ϵ-weakly universal sequence for Λ and A^ϵ which is 2ϵ-weakly universal for Λ and $\hat{A}_0$.

Example 6. Class of covariance stationary processes whose variances are uniformly bounded with $A_0 = \hat{A}_0$ = the real line and ρ the squared error distortion measure, [16, pp. 134-136].

Example 7. Transmission over a fixed-rate channel of capacity C. If we are to reproduce each source in the class with fidelity D at the output of the channel, the converse information transmission theorem requires $R_\theta(D) \le C$ for all $\theta \in \Lambda$ so that (4) is necessary.

In examples 3-7, the condition

$$R_\theta(D) < \infty \ \forall \ \theta \in \Lambda \tag{6}$$

appears either explicitly or implicitly. This condition is always necessary for Λ to be weakly universally encodable since otherwise no code exists which gives average distortion D for each member of the class. To replace (4)

of Theorem 2 by the considerably weaker requirement (6), we require that the following condition hold.

Condition (b).

(b.1) For each $\theta \in \Lambda$ there exists a real-valued function φ_θ defined on the integers such that $\lim_{n \to \infty} \varphi_\theta(n) = 0$ and such that for each positive integer n, each $F \in \mathcal{F}_{-\infty}^{0}$, and each $F_n \in \mathcal{F}_n^{\infty}$

$$|\mu_\theta(F \cap F_n) - \mu_\theta(F)\,\mu_\theta(F_n)| \le \varphi_\theta(n)\,\mu_\theta(F).$$

(b.2) The function φ_θ of (b.1) can be selected such that

$$\sum_{n=1}^{\infty} [\varphi_\theta(n)]^{\frac{1}{2}} < \infty, \ \forall\ \theta \in \Lambda.$$

(b.3) There is a $b \in \hat{A}_0$ for which

$$E_\theta[\rho(X_0, b)]^2 < \infty, \ \forall\ \theta \in \Lambda.$$

Remark. The following is equivalent to (b.1).

(b.1)' For each $\theta \in \Lambda$ there exists a real-valued function φ_θ defined on the integers such that $\lim_{n \to \infty} \varphi_\theta(n) = 0$ and such that for each positive integer n and each $E \in \mathcal{Q}_n^{\infty}$,

$$|\mu_\theta(E|\mathcal{F}_{-\infty}^{0}) - \mu_\theta(E)| \le \varphi_\theta(n)$$

with probability 1 (note that the conditional probability $\mu_\theta(E|\mathcal{F}_{-\infty}^{0})$ is a random variable). The equivalence of (b.1) and (b.1)' is demonstrated in [18]. If ρ is a metric on $A_0 \cup \hat{A}_0$, (b.3) is equivalent to the following condition.

(b.3)' For each $\theta \in \Lambda$ there exists a $b_\theta \in \hat{A}_0$ such that $E_\theta[\rho(X_0,b_\theta)]^2 < \infty$.

Examples of processes satisfying (b.1) and (b.2) are sequences of independent identically distributed random variables, sequences of m-dependent random variables [19], aperiodic Markov chains satisfying Doeblin's condition [20], [21] and stationary Gaussian processes with rational spectral density [21]-[24]. Chains of infinite order [25] satisfy (b.1). Note for example that the class of zero mean stationary Gaussian processes with finite average power and distortion measure $\rho(x,y) = |x-y|$ satisfies (b.3).

Theorem 5. Suppose A_0 is an abstract alphabet and $\hat{A}_0$ is countable. If Condition (b) is satisfied and if $R_\theta(D) < \infty$ for each $\theta \in \Lambda$, then the class is weakly universally encodable at distortion level D with reproduction alphabet $\hat{A}_0$.

Corollary 1. If in Theorem 5, $\hat{A}_0$ is given by either Example 1 or 2, then the conclusion is still true.

Finally we mention that Λ countable is also a sufficient condition to replace (4) of Theorem 2 by (6). This can be shown using techniques of the proof of Theorem 2.

Chapter 4. Strongly Universal Variable Rate Codes

In this chapter, we assume for convenience that $A_0 = \hat{A}_0$ and that the distortion measure ρ satisfies $\inf\{\rho(x,y) | y \in \hat{A}_0\} = 0$ for all $x \in A_0$.

__Definition 2.__ A family $\{(A,\mathcal{A},\mu_\theta | \theta \in \Lambda\}$ of stationary sources is strongly universally encodable at distortion level D if for each $\epsilon > 0$ there is an integer n_ϵ and a binary variable rate code on blocks of length n_ϵ, $\underline{B}_{n_\epsilon}$, such that for each $\theta \in \Lambda$,

$$\bar{\rho}_{n_\epsilon}(\underline{B}_{n_\epsilon},\theta) \leq D + \epsilon$$

and

$$\bar{r}_{n_\epsilon}(\underline{B}_{n_\epsilon},\theta) \leq (\ln 2)^{-1}[R_\theta(D) + \epsilon].$$

If strongly universal codes exist for distortion D for a family of sources with index set Λ, then there is a single binary variable rate code which has average distortion approximately D and average rate approximately $R_\theta(D)$ for each $\theta \in \Lambda$. Such codes are the variable-rate counterparts to strongly minimax universal fixed-rate codes [8].

Naturally, strongly universal codes do not exist for as broad a class of sources as weakly universal codes, and later in the section we give results that substantiate this. First, however, we give two sufficient conditions for strongly-universal codes to exist, the first involving the $\bar{\rho}$ distance measure between random processes [10].

__Theorem 6.__ Suppose A_0 is a finite alphabet with metric distortion measure ρ. Then any family of stationary ergodic sources $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ totally bounded under the $\bar{\rho}$ distance is strongly universally encodable for any nonnegative distortion level.

We now give another sufficient condition for the existence of strongly-universal codes. Following [26], we say a binary variable rate code (B_n, U_n, V_n) is a D-bounded distortion code if $\rho(\underline{x}^n, V_n[U_n(\underline{x}^n)]) \le D$ for all $\underline{x}^n \in A_0^n$. For integers $n \ge 1$, $\theta \in \Lambda$, and $D \ge 0$, define the OPTA function

$$r^b(n,D,\theta) = \inf\{\bar{r}_n(\underline{B}_n, \theta) \mid \underline{B}_n \text{ D-bounded}\}$$

This is well defined since $\underline{B}(\{\hat{A}_0^n\})$ is a D-bounded distortion code for all integers $n \ge 1$, $\theta \in \Lambda$, and $D \ge 0$. Define $r^b(D,\theta) = \inf\{r^b(n,D,\theta) \mid n \ge 1\}$. Standard subadditivity arguments can be used to show $r^b(D,\theta) = \lim_n r^b(n,D,\theta)$.

For any integer $n \ge 1$, we can form a binary variable rate code by combining the binary bit representation for a block code provided by Theorem 1 with the binary bit representation for $\hat{A}_0^n$, using the encoding rate of mapping a source word $\underline{x}$ into the shortest bit representation associated with a reproducing word $\underline{y}$ for which $\rho(\underline{x},\underline{y}) \le D$. In this way, we can prove Proposition 1.

<u>Proposition 1</u>. $r^b(D,\theta) = R_\theta(D)$ for $D > 0$, $(A,\mathcal{A},\mu_\theta)$ a stationary ergodic source, and $|A_0| < \infty$.

<u>Proposition 2</u>. Suppose A_0 is finite and $D > 0$. If $r^b(n,D,\theta)$ converges uniformly on Λ, then any family of stationary sources $\{(A,\mathcal{A},\mu_\theta) \mid \theta \in \Lambda\}$ for which $r^b(D,\theta) = R_\theta(D)$ for all $\theta \in \Lambda$ is strongly universally encodable for distortion level D.

The condition $r^b(D,\theta) = R_\theta(D)$, $\theta \in \Lambda$, is automatically fulfilled if $\{(A,\mathcal{A},\mu_\theta) \mid \theta \in \Lambda\}$ is a class of stationary ergodic sources. As an application of Proposition 2, we have

Theorem 7. The class $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ of all binary first-order Markov sources is strongly universally encodable for any nonnegative distortion level.

The following result prepares us to give partial converses to Theorems 6 and 7. Following this, we give two propositions which show that strongly-universal codes do not exist for a class of sources without restrictions on the memory and source alphabet of the class. The proof of Proposition 4 uses a construction contained in [8].

Proposition 3. Suppose A_0 is finite. If a family of stationary ergodic sources $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ is strongly universally encodable for positive distortion level D, then for any $\epsilon > 0$, there is a positive integer n depending on ϵ such that $R_{n,\theta}(D) < R_\theta(D) + \epsilon$ for all $\theta \in \Lambda$.

Proposition 4. Suppose A_0 is finite with metric distortion measure ρ. The class of all finite order Markov sources $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ is not strongly universally encodable for any distortion level D satisfying

$$0 < D < J^{-1}\min\{\sum_{y \in \hat{A}_0} \rho(x,y) | x \in A_0\}.$$

Proposition 5. Suppose A_0 is a metric space with unbounded metric distortion measure ρ. The class of all memoryless sources $\{(A,\mathcal{A},\mu_\theta) | \theta \in \Lambda\}$ is not strongly universally encodable for any nonnegative distortion level D.

Appendix A

Proof of Theorem 2. Let $D \geq 0$ satisfy $\sup\{R_\theta(D) \mid \theta \in \Lambda\} < \infty$. Choose $M > \sup\{(\ln 2)^{-1} R_\theta(D) + 2 \mid \theta \in \Lambda\}$, and let $\epsilon > 0$, $1 > \epsilon > 0$. For $\theta \in \Lambda$, define $\rho(\theta) = \max\{D+1, E_\theta\, \rho(X_0, b) + 1\}$.

Construction of Universal Codes

For each positive integer j, let $\hat{A}_0(j) = \{a_1, a_2, \ldots, a_j\} \cup \{b\}$, and let $\underline{b}^j$ be the j-tuple of b's. Define a subset K of the positive integers by $K = \{2^1, 2^2, 2^3, \ldots\}$. For each $k \in K$, let $n(k) = 2^{k \log_2 k}$, let $j_k = k^{-1} n(k)$, let $\mathcal{C}^*(k)$ be the collection of all block codes of length k, and define the set $\mathcal{C}(k) \subset \mathcal{C}^*(k)$ by

$$\mathcal{C}(k) = \{C \in \mathcal{C}^*(k) \mid C \subset [\hat{A}_0(\log_2 k)]^k,\ \underline{b}^k \in C, \text{ and } k^{-1}\lceil \log_2 |C| \rceil < M\} \tag{7}$$

Note that

$$|\mathcal{C}(k)| \leq \sum_{j=1}^{(\log_2 k)^k} \binom{(\log_2 k)^k}{j} \leq k^k,$$

so that we may index the elements of $\mathcal{C}(k)$ as $C(k,1), C(k,2), \ldots, C(k,v_k)$ where $v_k \leq k^k$ is a positive integer, and $|C(k,1| \leq |C(k,2)| \leq \ldots \leq |C(k,v_k)|$. For each $k \in K$ and each integer q, $1 \leq q \leq v_k$, define the random variable $f_{k,q} : A_0^{n(k)} \to [0,\infty)$ by

$$f_{k,q}(\underline{x}^{n(k)}) = j_k^{-1} \sum_{i=0}^{j_k - 1} \rho[(\underline{x}^{n(k)})_{i,k} \mid C(k,q)]$$

For each $k \in K$ and each $\underline{x}^{n(k)} \in A_0^{n(k)}$, define the set $\mathcal{C}(k, \underline{x}^{n(k)}) \subset \mathcal{C}(k)$ by

$$\mathcal{C}(k, \underline{x}^{n(k)}) = \{C(k,q) \in \mathcal{C}(k) \mid f_{k,q}(\underline{x}^{n(k)}) \leq D + \epsilon/2\} \tag{8}$$

For each $k \in K$, define a function $c_k : A_0^{n(k)} \to \mathcal{C}(k)$ as follows:
for $\underline{x}^{n(k)} \in A_0^{n(k)}$,

(i) if $\mathcal{C}(k,\underline{x}^{n(k)}) \neq \phi$, let $c_k(\underline{x}^{n(k)})$ be the code of least index in $\mathcal{C}(k,\underline{x}^{n(k)})$

(ii) if $\mathcal{C}(k,\underline{x}^{n(k)}) = \phi$, let $c_k(\underline{x}^{n(k)})$ be the code of least index in $\mathcal{C}(k)$ for which

$$j_k^{-1} \sum_{i=0}^{j_k-1} \rho[\underline{x}^{n(k)})_{i,k} | c_k(\underline{x}^{n(k)})] =$$

$$\min\{f_{k,q}(\underline{x}^{n(k)}) \mid 1 \le q \le v_k\}.$$

Note that c_k is well defined since $\mathcal{C}(k)$ is non empty for each k.

Claim 1. For $k \in K$, c_k is $\mathcal{A}_0^{n(k)}$ - measurable.

Proof. For $C(k,p) \in \mathcal{C}(k)$, $1 \le p \le v_k$, we have

$$\begin{aligned}
\{\underline{x}^{n(k)} | c_k(\underline{x}^{n(k)}) = C(k,p)\} &= \{\underline{x}^{n(k)} | \mathcal{C}(k,\underline{x}^{n(k)}) \neq \phi,\ c_k(\underline{x}^{n(k)}) = C(k,p)\} \\
&\cup \{\underline{x}^{n(k)} | \mathcal{C}(k,\underline{x}^{n(k)}) = \phi,\ c_k(\underline{x}^{n(k)}) = C(k,p)\} \\
&= (\{f_{k,1} > D+\epsilon/2\} \cap \{f_{k,2} > D+\epsilon/2\} \cap \ldots \cap \{f_{k,p-1} > D+\epsilon/2\} \\
&\cap \{f_{k,p} \le D+\epsilon/2\}) \cup (\{f_{k,1} > D+\epsilon/2\} \cap \ldots \cap \{f_{k,v_k} > D+\epsilon/2\} \\
&\cap \{f_{k,p} = \min\{f_{k,q} | 1 \le q \le v_k\}\} \cap \{f_{k,p} < \min\{f_{k,q} | 1 \le q \le p-1\}\}) \\
&\in \mathcal{A}_0^{n(k)}
\end{aligned}$$

For each $k \in K$, define $h_k : A \to [0,\infty)$ by

$$h_k(\omega) = j_k^{-1} \sum_{i=0}^{j_k-1} \rho[\underline{X}^k(T^{ik}\omega)|c_k(\underline{X}^{n(k)}(\omega))]$$

Claim 2. For $k \in K$, h_k is $\mathcal{A}$-measurable.

Proof. For α a real number

$$\{\omega|h_k(\omega)<\alpha\} = \bigcup_{1\le q\le v_k} \{\omega|c_k(\underline{X}^{n(k)}(\omega)) = C(k,q), f_{k,q}(\underline{X}^{n(k)}(\omega)) < \alpha\} \in \mathcal{A}.$$

For each $k \in K$ and each integer i, $1 \le i \le v_k$, let $w_{k,i}$ be the $(\log_2 k)^k$ bit binary representation of i, and let $(B_k^i, U_k^i, V_k^i) = \underline{B}(C(k,i))$. For each $k \in K$, define the code $\underline{B}_{n(k)} = (B_{n(k)}, U_{n(k)}, V_{n(k)})$ by

(a) $B_{n(k)} = \bigcup_{1 \le i \le v_k} (w_{k,i} \times [B_k^i]^{j_k})$

(b) $U_{n(k)}(\underline{x}^{n(k)}) = (w_{k,i}, U_k^i((\underline{x}^{n(k)})_{0,k}), \ldots, U_k^i((\underline{x}^{n(k)})_{j_k-1,k}))$,

$\underline{x}^{n(k)} \in \{\underline{x} \in A_0^{n(k)}|c_k(\underline{x}) = C(k,i)\}, i=1,\ldots,v_k$.

(c) $V_{n(k)}(w_{k,i},\underline{b}) = (V_k^i(b_0), V_k^i(b_1), \ldots, V_k^i(b_{j_k-1}))$,

$\underline{b} = (b_0, b_1, \ldots, b_{j_k-1}) \in [B_k^i]^{j_k}, i=1,\ldots,v_k$.

Claim 3. For $k \in K$, $U_{n(k)}$ is $\mathcal{A}^{n(k)}$-measurable.

Proof. $\{\underline{x}^{n(k)}|U_{n(k)}(\underline{x}^{n(k)}) = (w_{k,i}, b_0, b_1, \ldots, b_{j_k-1})\} = \{\underline{x}^{n(k)}|c_k(\underline{x}^{n(k)}) = C(k,i)\}$

$\cap (\bigcap_{0 \le r \le j_k-1} \{\underline{x}^{n(k)}|U_k^i((\underline{x}^{n(k)})_{r,k}) = b_r\}) \in \mathcal{A}_0^{n(k)}$

For each $k \in K$, since $B_{n(k)}$ is a finite set of binary sequences satisfying the prefix condition, $U_{n(k)}: A_0^{n(k)} \to B_{n(k)}$ is $\mathcal{A}_0^{n(k)}$-measurable, and $V_{n(k)}: B_{n(k)} \to \hat{A}_0^{n(k)}$,

$\underline{B}_{n(k)} = (B_{n(k)}, U_{n(k)}, V_{n(k)})$ is a binary variable rate code on blocks of length $n(k)$.

Note that from (1), for each $k \in K$ and each $\omega \in A$,

$$\rho(\underline{X}^{n(k)}(\omega), V_{n(k)}[U_{n(k)}(\underline{X}^{n(k)}(\omega))]) = h_k(\omega) \tag{9}$$

and from (2),

$$\ell[U_{n(k)}(\underline{X}^{n(k)}(\omega))] = k^{-1} n(k) \lceil \log_2 |c_k(\underline{X}^{n(k)}(\omega))| \rceil + (\log_2 k)^k \tag{10}$$

Average Distortion of Codes

Assume for the moment the truth of the following lemma.

Lemma 1. For each $\theta \in \Lambda$, there is an integer $r(\theta) \in K$ and a code $C(\theta) \in \mathcal{C}^*(r(\theta))$ containing $\underline{b}^{r(\theta)}$ and satisfying

$$\mu_\theta\{j^{-1} \sum_{i=0}^{j-1} \rho[\underline{X}^{r(\theta)} \circ T^{ir(\theta)} | C(\theta)] > D + (\epsilon/2)\} < \delta^*(\theta) \tag{11}$$

for each positive integer j and

$$\rho(\theta)\, \delta^*(\theta) < \epsilon/4 \tag{12}$$

$$M\, \delta^*(\theta) < \epsilon/4 \ln 2 \tag{13}$$

$$r(\theta)^{-1} \lceil \log_2 |C(\theta)| \rceil < (\ln 2)^{-1} [R_\theta(D) + (\epsilon/2)] \tag{14}$$

In the remainder of the proof, we fix $\theta \in \Lambda$. For $r(\theta)$ given by Lemma 1, and integers $k \geq r(\theta)$, $k \in K$, define the random variable $g_k : A \to [0,\infty)$ by

$$g_k(\omega) = j_k^{-1} \sum_{i=0}^{j_k-1} \rho[\underline{X}^k(T^{ik}\omega) | [C(\theta)]^{k/r(\theta)}]$$

Note that

$$g_k(\omega) = r(\theta)^{-1}n(k) \sum_{i=0}^{r(\theta)^{-1}n(k)-1} \rho[\underline{X}^{r(\theta)}(T^{ir(\theta)}\omega)\,|C(\theta)]$$

and therefore from Lemma 1,

$$\mu_\theta\{g_k > D + (\epsilon/2)\} < \delta^*(\theta) \tag{15}$$

For $k \geq r(\theta)$, $k \in K$, from (14)

$$\log_2|[C(\theta)]^{k/r(\theta)}| = \log_2|C(\theta)|^{k/r(\theta)}$$

$$< r(\theta)^{-1}k\{r(\theta)(\ln 2)^{-1}[R_\theta(D) + (\epsilon/2)] + 1\}$$

$$< k((\ln 2)^{-1}R_\theta(D) + 2) \tag{16}$$

It follows that for all large k, $k \in K$,

$$k^{-1}\lceil \log_2|[C(\theta)]^{k/r(\theta)}| \rceil < M. \tag{17}$$

Now choose $k(\theta) \geq r(\theta)$, $k(\theta) \in K$, such that $k \geq k(\theta)$, $k \in K$ implies inequality (17) is satisfied and $C(\theta) \subset [\hat{A}_0(\log_2 k)]^{r(\theta)}$. Then for $k \geq k(\theta)$, $k \in K$,

$$[C(\theta)]^{k/r(\theta)} \subset [\hat{A}_0(\log_2 k)]^k \tag{18}$$

and since $\underline{b}^{r(\theta)} \in C(\theta)$,

$$\underline{b}^k \in [C(\theta)]^{k/r(\theta)} \tag{19}$$

For $k \geq k(\theta)$, $k \in K$, (17) through (19) imply

$$[C(\theta)]^{k/r(\theta)} \in \mathcal{C}(k) \tag{20}$$

We now establish three claims.

Claim 4. For $k \in K$, $h_k(\omega) > D + (\epsilon/2)$ implies $\mathcal{C}(k, \underline{X}^{n(k)}(\omega)) = \phi$.

Proof. If $\mathcal{C}(k, \underline{X}^{n(k)}(\omega)) \neq \phi$, then by (i), $h_k(\omega) \le D + (\epsilon/2)$, a contradiction.

Claim 5. For $k \ge k(\theta)$, $k \in K$, $h_k(\omega) > D + (\epsilon/2)$ implies $g_k(\omega) > D + (\epsilon/2)$.

Proof. By Claim 4, $\mathcal{C}(k, \underline{X}^{n(k)}(\omega)) = \phi$. Since $[C(\theta)]^{k/r(\theta)} \in \mathcal{C}(k)$ by (20), (8) implies $g_k(\omega) > D + (\epsilon/2)$.

Claim 6. For $k \ge k(\theta)$, $k \in K$, $h_k \le g_k$ on $\{h_k > D + (\epsilon/2)\}$

Proof. By Claim 4, $h_k(\omega) > D + (\epsilon/2)$ implies $\mathcal{C}(k, \underline{X}^{n(k)}(\omega)) = \phi$. Since $[C(\theta)]^{k/r(\theta)} \in \mathcal{C}(k)$, (ii) implies $h_k(\omega) \le g_k(\omega)$.

Note that Claims 5 and 6 imply

$$E_\theta h_k \chi_{\{h_k > D + (\epsilon/2)\}} \le E_\theta g_k \chi_{\{g_k > D + (\epsilon/2)\}} \tag{21}$$

for $k \ge k(\theta)$, $k \in K$, where χ_F is the indicator function for the set F, $F \in \mathcal{A}$.

Finally, using (9) and (21), we have for $\underline{B}_{n(k)}$ for $k \ge k(\theta)$, $k \in K$,

$$\bar{\rho}_{n(k)}(\theta) = E_\theta \rho(\underline{X}^{n(k)}, V_{n(k)}[U_{n(k)}(\underline{X}^{n(k)})])$$

$$= E_\theta h_k(\chi_{\{h_k \le D + (\epsilon/2)\}} + \chi_{\{h_k > D + (\epsilon/2)\}})$$

$$\le D + (\epsilon/2) + E_\theta g_k \chi_{\{g_k > D + (\epsilon/2)\}} \tag{22}$$

Note that since $\underline{b}^{r(\theta)} \in C(\theta)$, $g_k \le \rho(\underline{X}^{n(k)}, \underline{b}^{n(k)})$ and thus

$$E_\theta g_k \chi_{\{g_k > D + (\epsilon/2)\}} \le E_\theta \rho(\underline{X}^{n(k)}, \underline{b}^{n(k)}) \chi_{\{g_k > D + (\epsilon/2)\}}$$

$$\le \rho(\theta)\mu_\theta\{g_k > D + (\epsilon/2)\} + E_\theta \rho(\underline{X}^{n(k)}, \underline{b}^{n(k)}) \chi_{\{\rho(\underline{X}^{n(k)}, \underline{b}^{n(k)}) > \rho(\theta)\}} \tag{23}$$

Call the second term on the right side of the above inequality $q(n(k),\theta)$. Combining (12), (15), (22), and (23) gives

$$\bar{\rho}_{n(k)}(\theta) < D + (3\epsilon/4) + q(n(k),\theta) \tag{24}$$

for $k \geq k(\theta)$, $k \in K$. Under the conditions of the theorem, Berger [16] has shown $q(n(k),\theta) \underset{k}{\rightarrow} 0$, $k \in K$.

<u>Average Rate of Codes</u>

For $k \geq k(\theta)$, $k \in K$, and $\omega \in \{g_k \leq D + (\epsilon/2)\}$, (8) and (20) imply

$$[C(\theta)]^{k/r(\theta)} \in \mathcal{C}(k, \underline{X}^{n(k)}(\omega)).$$

Thus from (i),

$$k^{-1}\lceil \log_2 |c_k(\underline{X}^{n(k)}(\omega))| \rceil \leq k^{-1}\lceil \log_2 |[C(\theta)]^{k/r(\theta)}| \rceil$$

$$\leq r(\theta)^{-1}\lceil \log_2 |C(\theta)| \rceil + k^{-1} . \tag{25}$$

Using (10) and (25) gives

$$\ell[U_{n(k)}(\underline{X}^{n(k)}(\omega))] \leq r(\theta)^{-1} n(k) \lceil \log_2 |C(\theta)| \rceil + k^{-1} n(k) + (\log_2 k)^k \tag{26}$$

Then for $k \geq k(\theta)$, $k \in K$, using (7), (10), and (26), we have for $\underline{B}_{n(k)}$,

$$\bar{r}_{n(k)}(\theta) = E_\theta n(k)^{-1} \ell[U_{n(k)}(\underline{X}^{n(k)})](\chi_{\{g_k \leq D+(\epsilon/2)\}} + \chi_{\{g_k > D+(\epsilon/2)\}})$$

$$\leq r(\theta)^{-1}\lceil \log_2 |C(\theta)| \rceil + k^{-1} + M\mu_\theta\{g_k > D + (\epsilon/2)\} + n(k)^{-1}(\log_2 k)^k \tag{27}$$

Using (13) through (15) gives

$$\bar{r}_{n(k)}(\theta) < (\ln 2)^{-1}[R_\theta(D) + (3\epsilon/4)] + k^{-1} + n(k)^{-1}(\log_2 k)^k \tag{28}$$

for $k \geq k(\theta)$, $k \in K$. The last two terms decrease to 0 with k, $k \in K$.

Construction of Sequence of Codes

From this sequence of codes $\{\underline{B}_{n(k)}\}_{k \in K}$, we can construct a sequence of binary variable rate codes on blocks of lengths 1, 2, 3,... which will have the asymptotic properties of the original codes. It suffices to show how to construct a binary variable rate code on blocks of length n for an integer n satisfying $n(k) \le n \le n(k+1)$. We follow the construction of [12, Appendix A] first using $\underline{B}_{n(k)}$ as many times as possible on the source word $\underline{x}^n$ of length n, then using $\underline{B}_{n(k-1)}$ as many times as possible on the remaining n mod n(k) letters, and continuing in this fashion until the number $\underline{m}$ of remaining letters is less than n(1). As in [12], this defines a binary variable rate code on blocks of length n, $\underline{B}_n$, where the encoding mapping U_n ignores the $\underline{m}$ remaining letters and the decoding mapping V_n adds a sequence of $\underline{m}$ b's, to give a full n-tuple of reproducing letters.

If we let $a_{k+1}(n) = n$, $a_0(n) = 0$, and $a_i(n) = a_{i+1}(n) \bmod n(i)$ for $1 \le i \le k$, we see that $a_{i+1}(n) - a_i(n)$ is the number of letters of $\underline{x}^n$ encoded using $\underline{B}_{n(i)}$ for $1 \le i \le k$, and it is the number $\underline{m}$ of letters ignored by U_n for $i = 0$. The average distortion of $\underline{B}_n$ is

$$\bar{\rho}_n(\theta) = n^{-1}\{\sum_{i=1}^{k} [a_{i+1}(n) - a_i(n)]\bar{\rho}_{n(i)}(\theta) + [a_1(n) - a_0(n)]E_\theta\, \rho(X_0, b)\} \quad (29)$$

and the average rate is

$$\bar{r}_n(\theta) = n^{-1} \sum_{i=1}^{k} [a_{i+1}(n) - a_i(n)]\bar{r}_{n(i)}(\theta) \quad (30)$$

It follows from (24) and (28) that there is an integer $m(\theta) \in K$ such that for $m \ge m(\theta)$, $m \in K$,

$$\bar{\rho}_{n(m)}(\theta) < D + (7\epsilon/8)$$

and $$\bar{r}_{n(m)}(\theta) < (\ln 2)^{-1}\,[R_\theta(D) + (7\epsilon/8)]\ .$$

For $i = 1,\ldots,m(\theta) - 1$, from (7) and the code construction

$$\bar{\rho}_{n(i)}(\theta) \leq E_\theta\, \rho(X_0,b)$$

and $$\bar{r}_{n(i)}(\theta) \leq M + n(i)^{-1}(\log_2 i)^i. \tag{31}$$

Then for $n \geq n(m(\theta))$ we have from (29) and (30) that

$$\bar{\rho}_n(\theta) < n^{-1}\,[n(D + (7\epsilon/8)) + n(m(\theta))E_\theta\rho(X_0,b)]$$

and

$$\bar{r}_n(\theta) < n^{-1}[n(\ln 2)^{-1}\,[R_\theta(D) + (7\epsilon/8)] + n(m(\theta))\ \max\{\bar{r}_{n(i)}(\theta)\,|\,1 \leq i \leq m(\theta)-1\}]$$

It follows that there is an integer $N(\theta)$ such that for $n \geq N(\theta)$,

$$\bar{\rho}_n(\theta) \leq D + \epsilon \tag{32}$$

and $$\bar{r}_n(\theta) \leq (\ln 2)^{-1}\,[R_\theta(D) + \epsilon]\ . \tag{33}$$

It only remains to prove Lemma 1 since (32) and (33) establish the theorem.

<u>Proof of Lemma 1</u>. In this proof, we fix $\theta \in \Lambda$. All chosen constants and defined functions depend on the particular $\theta \in \Lambda$ fixed, but for convenience, this is not explicitly denoted. Choose $a > 1$ such that

$$(a-1)^{-1}(\rho(\theta))^2 < 8^{-1}\epsilon \tag{34}$$

$$(a-1)^{-1}\rho(\theta)\,M < (8\ln 2)^{-1}\epsilon \tag{35}$$

Choose $\delta > 0$ small enough such that $\delta < 1$ and

$$2a\delta < \epsilon/2 \tag{36}$$

Choose an integer $s \geq 0$ such that for integers $t \geq s$,

$$(D+a\delta)^{-1}\rho(\theta)q(t,\theta) < 8^{-1}\epsilon \tag{37}$$

$$(D+a\delta)^{-1}M\,q(t,\theta) < (8\ln 2)^{-1}\epsilon \tag{38}$$

Since $E_\theta\,\rho(X_0,b) < \infty$ and $R_\theta(D) < \infty$, we can apply Theorem 1 to obtain for an integer $r \geq s$, $r \in K$, a code $C \in \mathcal{C}^*(r)$ containing $\underline{b}^r$ and satisfying (14) and

$$\mu_\theta\{\rho(\underline{X}^r|C) > D+\delta\} < \delta \tag{39}$$

We now show that we can find a constant δ^* which satisfies (11) through (13).

Define $S \in \mathcal{A}$ by $S = \{\rho(\underline{X}^r|C) \in (D+\delta, D+a\delta)\}$, and define $T \in \mathcal{A}$ by $T = \{\rho(\underline{X}^r|C) > \rho(\theta)\}$. Define random variables $p : A \to [0,\infty)$ and $q : A \to [0,\infty)$ by

$$p = (D+\delta)\chi_S + \rho(\underline{X}^r|C)\chi_{S^c}$$

and

$$q = p\chi_{\{p \leq \rho(\theta)\}}\,.$$

Note that

$$q \leq p \leq \rho(\underline{X}^r|C) < p + a\delta \tag{40}$$

and

$$p = q + \rho(\underline{X}^r|C)\chi_T \tag{41}$$

Then for each positive integer j, using (36), (40), and (41),

$$\mu_\theta\{j^{-1}\sum_{i=0}^{j-1}\rho(\underline{X}^r \circ T^{ir}|C) > D+(\epsilon/2)\} \leq \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}\rho(\underline{X}^r \circ T^{ir}|C) > D+2a\delta\}$$

$$\leq \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}(p \circ T^{ir}) > D+a\delta\}$$

$$= \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}[(q \circ T^{ir}) + \rho(\underline{X}^r \circ T^{ir}|C)(\chi_T \circ T^{ir})] > D+a\delta\}$$

$$\leq \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}(q\circ T^{ir}) > D+a\delta\} + \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}\rho(\underline{X}^r\circ T^{ir}|C)(\chi_T\circ T^{ir}) > D+a\delta\} \tag{42}$$

where the last inequality follows from the union bound. Let I_1 and I_2 denote the first and second terms given by the union bound. From the Chebychev inequality,

$$\begin{aligned} I_2 &\leq (D+a\delta)^{-1}E_\theta\, j^{-1}\sum_{i=0}^{j-1}\rho(\underline{X}^r\circ T^{ir}|C)(\chi_T\circ T^{ir}) \\ &= (D+a\delta)^{-1}E_\theta\,\rho(\underline{X}^r|C)\chi_T \\ &\leq (D+a\delta)^{-1}E_\theta\,\rho(\underline{X}^r,\underline{b}^r)\chi_{\{\rho(\underline{X}^r,\underline{b}^r)>\rho(\theta)\}} \\ &= (D+a\delta)^{-1}q(r,\theta) \end{aligned} \tag{43}$$

We now consider I_1. Note that

$$\sum_{i=0}^{j-1}(\chi_{\{q\leq D+\delta\}}\circ T^{ir}) + \sum_{i=0}^{j-1}(\chi_{\{q\geq D+a\delta\}}\circ T^{ir}) = j \tag{44}$$

and

$$\sum_{i=0}^{j-1}(q\circ T^{ir}) \leq (D+\delta)\sum_{i=0}^{j-1}(\chi_{\{q\leq D+\delta\}}\circ T^{ir}) + \rho(\theta)\sum_{i=0}^{j-1}(\chi_{\{q\geq D+a\delta\}}\circ T^{ir}) \tag{45}$$

For ω such that

$$\sum_{i=0}^{j-1}q(T^{ir}\omega) > j(D+a\delta),$$

(45) gives

$$(D+\delta)\sum_{i=0}^{j-1}\chi_{\{q\leq D+\delta\}}(T^{ir}\omega) + \rho(\theta)\sum_{i=0}^{j-1}\chi_{\{q\geq D+a\delta\}}(T^{ir}\omega) > j(D+a\delta),$$

and further algebra using (44) reduces this to

$$j^{-1}\sum_{i=0}^{j-1}\chi_{\{q\geq D+a\delta\}}(T^{ir}\omega) > (\rho(\theta)-(D+\delta))^{-1}(a-1)\delta.$$

Thus, we have

$$\{j^{-1}\sum_{i=0}^{j-1}(q\circ T^{ir}) > D+a\delta\} \subset \{j^{-1}\sum_{i=0}^{j-1}(\chi_{\{q\geq D+a\delta\}}\circ T^{ir}) > (\rho(\theta)-(D+\delta))^{-1}(a-1)\delta\} \tag{46}$$

Then using (46),

$$I_1 \leq \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}(\chi_{\{q\geq D+a\delta\}}\circ T^{ir}) > (\rho(\theta)-(D+\delta))^{-1}(a-1)\delta\} .$$

Since $(\rho(\theta)-(D+\delta))^{-1}(a-1)\delta > 0$, we can apply the Chebychev inequality to the above term and then use (40) and (39) to obtain

$$\begin{aligned} I_1 &\leq ((a-1)\delta)^{-1}(\rho(\theta)-(D+\delta))\, E_\theta\, j^{-1}\sum_{i=0}^{j-1}(\chi_{\{q\geq D+a\delta\}}\circ T^{ir}) \\ &< ((a-1)\delta)^{-1}\rho(\theta)\, E_\theta \chi_{\{q\geq D+a\delta\}} \\ &\leq ((a-1)\delta)^{-1}\rho(\theta)\, E_\theta \chi_{\{\rho(\underline{X}^r|C)\geq D+a\delta\}} \\ &< (a-1)^{-1}\rho(\theta) \end{aligned} \tag{47}$$

Then from (42), (43), and (47), for each positive integer j,

$$\mu_\theta\{j^{-1}\sum_{i=0}^{j-1}\rho(\underline{X}^r\circ T^{ir}|C) > D+(\epsilon/2)\} < (a-1)^{-1}\rho(\theta) + (D+a\delta)^{-1}q(r,\theta)$$

Defining δ^* to be the right hand side of the above inequality, we note that (11) is established and using (34), (35), (37), and (38), we have

$$\rho(\theta)\delta^* < 4^{-1}\epsilon$$

$$M\delta^* < (4\ln 2)^{-1}\epsilon$$

establishing (12) and (13).

Proof of Theorem 3.

Let $R'_\theta(\cdot)$ be the rate-distortion function for source $(A,\mathcal{A},\mu_\theta)$ and reproducing alphabet A'_0, $\theta \in \Lambda$. We have from Condition (a) that $R'_\theta(D+\varepsilon) \le R_\theta(D)$ for each $\varepsilon > 0$. Since $R'_\theta(\cdot)$ is a continuous function, this implies

$$R'_\theta(D) \le R_\theta(D) \ . \tag{48}$$

Condition (a) also implies that for $b \in \hat{A}_0$, there is a $b' \in A'_0$ for which

$$E_\theta\, \rho(X_0,b') \le E_\theta\, \rho(X_0,b) + 1 < \infty \tag{49}$$

Equations (48) and (49) enable us to apply Theorem 2 to obtain for $\varepsilon > 0$, a sequence of codes $\{\underline{B}_n\}$ and for each $\theta \in \Lambda$ an integer $N(\theta)$ such that for $n \ge N(\theta)$,

$$\bar{\rho}_n\,(\theta) \le D + \varepsilon$$

and

$$\bar{r}_n\,(\theta) \le (\ln 2)^{-1}[R'_\theta(D) + \varepsilon].$$

Applying (48) gives the result of the theorem.

Proof of Theorem 4. Let $\varepsilon > 0$. As discussed in [14], we can assume in proving the theorem that $R_\theta(D) < \infty$ for all $\theta \in \Lambda$. This condition and Condition (a) imply that the map $\theta \to R_\theta(D)$ is $\mathcal{L}$-measurable [14]. Since the sets $\Lambda_n = \{\theta \in \Lambda | R_\theta(D) \le n\} \in \mathcal{L}$ increase to Λ, we can find a set $\Lambda_N \in \mathcal{L}$ with $W(\Lambda_N) \ge 1 - (\varepsilon/2)$ for which $\theta \in \Lambda_N$ implies

$$\sup\,\{R_\theta(D) | \theta \in \Lambda_N\} \le N \tag{50}$$

and

$$E_\theta\, \rho\,(X_0,b) < \infty \tag{51}$$

Using (50) and (51) to apply Theorem 3, we obtain a sequence of codes $\{\underline{B}_n\}$ and for each $\theta \in \Lambda_N$ an integer $N'(\theta)$ for which $n \ge N'(\theta)$ implies

$$\bar{\rho}_n(\theta) \le D + (\varepsilon/2)$$

and

$$\bar{r}_n(\theta) \le (\ln 2)^{-1}[R_\theta(D) + (\epsilon/2)].$$

For each n, define functions $f_n:\Lambda_N \to [0,\infty)$ and $g_n:\Lambda_N \to [0,\infty)$ by

$$f_n(\theta) = \sup\{\bar{\rho}_k(\theta) | k \ge n\}$$

and

$$g_n(\theta) = \sup\{\bar{r}_k(\theta) | k \ge n\}.$$

We have that f_n and g_n are $\{B \cap \Lambda_N | B \in \mathcal{L}\}$-measurable functions and

$$\lim_n f_n(\theta) \le D + (\epsilon/2),$$

$$\lim_n g_n(\theta) \le (\ln 2)^{-1}[R_\theta(D) + (\epsilon/2)],$$

for each $\theta \in \Lambda_N$. From Egoroff's theorem, we can obtain a set $\Lambda' \in \mathcal{L}$ with $W(\Lambda') \ge 1 - \epsilon$ and an integer N^* such that for $\theta \in \Lambda'$ and $n \ge N^*$,

$$f_n(\theta) \le D + \epsilon$$

and

$$g_n(\theta) \le (\ln 2)^{-1}[R_\theta(D) + \epsilon].$$

This establishes the theorem.

<u>Proof of Theorem 5.</u> With the exception of an analogue to Lemma 1, the proof of Theorem 5 is very similar to the proof of Theorem 2. Therefore, we only indicate necessary changes to the proof of Theorem 2.

First let $D \ge 0$ satisfy $R_\theta(D) < \infty$ for all $\theta \in \Lambda$. In place of (7), we make the definition

$$\mathcal{C}(k) = \{C \in \mathcal{C}^*(k) | C \subset [\hat{A}_0(\log_2 k)]^k, \underline{b}^k \in C, \text{ and } |C| \le j_k\} \qquad (52)$$

In place of Lemma 1, we have

<u>Lemma 2.</u> *For each $\theta \in \Lambda$, there is an integer $r(\theta) \in K$ and a code $C(\theta)$ $\in \mathcal{C}^*(r(\theta))$ containing $\underline{b}^{r(\theta)}$ and satisfying*

$$\mu_\theta\{j^{-1}\sum_{i=0}^{j-1} \rho(\underline{X}^{r(\theta)} | C(\theta)) > D + (\epsilon/2)\} \le (j\,\epsilon^2)^{-1}\, 16 a_\theta \qquad (53)$$

for each positive integer j where $a_\theta < \infty$ and

$$r(\theta)^{-1} \lceil \log_2 |C(\theta)| \rceil < (\ln 2)^{-1} [R_\theta(D) + (\epsilon/2)]. \tag{54}$$

Using (53), we have corresponding to (15)

$$\mu_\theta\{g_k > D + (\epsilon/2)\} \leq 16(n_k \epsilon^2)^{-1} r(\theta) a_\theta. \tag{55}$$

From (16), we have for all large $k \in K$,

$$\log_2 |[C(\theta)]^{k/r(\theta)}| < k \log_2 k - \log_2 k,$$

implying

$$|[C(\theta)]^{k/r(\theta)}| < k^{-1} 2^{k \log_2 k} = j_k,$$

which replaces (17). As in the proof of Theorem 2, we can then show there is an integer $k(\theta)$, $k(\theta) \in K$, for which $k \geq k(\theta)$, $k \in K$, implies $[C(\theta)]^{k/r(\theta)} \in \mathcal{C}(k)$. Then from (22), (23), and (55), we have

$$\bar{\rho}_{n(k)}(\theta) \leq D + (\epsilon/2) + 16(n_k \epsilon^2)^{-1} \rho(\theta) r(\theta) a_\theta + q(n(k),\theta) \tag{56}$$

for $k \geq k(\theta)$, $k \in K$.

We now examine the average rate of $\underline{B}_{n(k)}$. From (10), (26), and (52),

$$\bar{r}_{n(k)}(\theta) \leq r(\theta)^{-1} \lceil \log_2 |C(\theta)| \rceil + k^{-1} + k^{-1} \lceil \log_2 j_k \rceil \mu_\theta\{g_k > D + (\epsilon/2)\}$$
$$+ n(k)^{-1} (\log_2 k)^k$$

for $k \geq k(\theta)$, $k \in K$, corresponding to (27). Substituting (54) and (55) in the above equation gives

$$\bar{r}_{n(k)}(\theta) \leq (\ln 2)^{-1} [R_\theta(D) + (\epsilon/2)] + 16(\epsilon k)^{-2} \{ j_k^{-1} \lceil \log_2 j_k \rceil \} r(\theta) a_\theta$$
$$+ k^{-1} + n(k)^{-1} (\log_2 k)^k \tag{57}$$

for $k \geq k(\theta)$, $k \in K$. The last three terms decrease to 0 with k, $k \in K$ ($j_k \underset{k}{\to} \infty$).

To construct a sequence of binary variable rate codes on blocks of length 1,2,3,..., we follow the argument of Theorem 2, using (56) and (57) instead of (24) and (28) and using (10) and (52) to replace (31) with

$$\bar{r}_{n(i)}(\theta) \le 2^{-i}\lceil \log_2 j_i \rceil + n(i)^{-1}(\log_2 i)^i$$

for $i=1,\ldots,m(\theta)-1$. It only remains to prove Lemma 2, but first we prove a useful lemma.

<u>Lemma 3</u>. Fix $\theta \in \Lambda$. For q a positive integer, let g be an A_0^q-measurable function such that for $f = g \circ \underline{X}^q$, $E_\theta f=0$ and $E_\theta|f|^2 < \infty$. Then if (b.1) and (b.2) of Condition (b) hold,

$$E_\theta\Big[\sum_{i=0}^{j-1} f \circ T^{iq}\Big]^2 \le a_\theta j$$

for each positive integer j where $a_\theta < \infty$.

<u>Proof of Lemma 3</u>. Let j be a positive integer.

$$\begin{aligned}
E_\theta\Big[\sum_{i=0}^{j-1} f \circ T^{iq}\Big]^2 &= E_\theta\Big[\sum_{i=0}^{j-1}[f\circ T^{iq}]^2 + 2\sum_{0\le m<n\le j-1}(f\circ T^{mq})(f\circ T^{nq})\Big] \\
&\le \sum_{i=0}^{j-1} E_\theta[f\circ T^{iq}]^2 + 2\sum_{0\le m<n\le j-1}|E_\theta(f\circ T^{mq})(f\circ T^{nq})| \\
&\le j\, E_\theta f^2 + 2j\sum_{k=1}^{j-1}|E_\theta[f][f\circ T^{kq}]| \\
&\le j\{E_\theta f^2 + 2\sum_{k=1}^{\infty}|E_\theta[f][f\circ T^{kq}]|\} \qquad (58)
\end{aligned}$$

Under the hypothesis of the lemma, it follows from [18, Lemma 1.1] that

$$|E_\theta[f][f\circ T^{kq}]| \le 2[\varphi_\theta(kq)]^{\frac{1}{2}} E_\theta|f|^2$$

for $k=1,2,\ldots$ Then using (58),

$$E_\theta\Big[\sum_{i=0}^{j-1} f\circ T^{iq}\Big]^2 \le j\{E_\theta f^2 + 4E_\theta|f|^2\sum_{k=1}^{\infty}[\varphi_\theta(kq)]^{\frac{1}{2}}\}$$

establishing the lemma.

Proof of Lemma 2. Fix $\theta \in \Lambda$. Some chosen constants and defined functions depend on the particular $\theta \in \Lambda$ fixed, but for convenience, this is not explicitly denoted. Since by hypothesis, $E_\theta\ \rho\ (X_0,b) < \infty$ and $R_\theta(D) < \infty$, we can apply Theorem 1 to obtain an integer $r \in K$ and a code $C \in \mathcal{C}^*(r)$ containing $\underline{b}^r$ and satisfying (54) and

$$E_\theta\ \rho\ (\underline{X}^r|c) = D + \delta \tag{59}$$

where $\varepsilon/4 > \delta \geq 0$. Define a function f: $A \to [o,\infty)$ by $f(\omega) = \rho(\underline{X}^r(\omega)|C) - (D + \delta)$. Then

$$E_\theta\ f = 0 \tag{60}$$

and for $p > 1$,

$$\begin{aligned} E_\theta|f|^p &\leq E_\theta[|\rho(\underline{X}^r|C)| + |D+\delta|]^p \\ &\leq 2^{p-1}[E_\theta[\rho(\underline{X}^r|C)]^p + (D+\delta)^p] \\ &\leq 2^{p-1}[E_\theta[\rho(\underline{X}^r,\underline{b}^r)]^p + (D+\delta)^p] \\ &\leq 2^{p-1}[E_\theta[\rho(X_0,b)]^p + (D+\delta)^p]. \end{aligned} \tag{61}$$

Since Condition (b) holds, (61) gives

$$E_\theta|f|^2 < \infty \tag{62}$$

Using (60) and (62) to apply Lemma 3 gives

$$E_\theta[\sum_{i=0}^{j-1} f\circ T^{ir}]^2 \leq a_\theta j \tag{63}$$

for each positive integer j where $a_\theta < \infty$.

For each positive integer j

$$\mu_\theta\{j^{-1}\sum_{i=0}^{j-1}\rho\,(\underline{X}^r\bullet T^{ir}|C) > D + (\epsilon/2)\}$$

$$= \mu_\theta\{j^{-1}\sum_{i=0}^{j-1}[\rho(\underline{X}^r\bullet T^{ir}|C) - (D+\delta)] > [(\epsilon/2) - \delta]\}$$

$$= \mu_\theta\ \{\sum_{i=0}^{j-1}(f\bullet T^{ir}) > j[(\epsilon/2) - \delta]\}$$

Applying Chebychev's inequality to the right hand term and using (63), we have for each positive integer j,

$$\mu_\theta\{j^{-1}\sum_{i=0}^{j-1}(\underline{X}^r\bullet T^{ir}|C) > D + (\epsilon/2)\} \le (j[(\epsilon/2) - \delta])^{-2}E_\theta[\sum_{i=0}^{j-1}(f\bullet T^{ir})]^2 \qquad (64)$$

$$\le (j(\epsilon/4))^{-2}\ a_\theta\, j \qquad (65)$$

where $a_\theta < \infty$, which establishes (53).

Proof of Corollary 1. In the case of Example 1 or 2, Kieffer shows that the A_0' of Condition (a) can be taken as a countable dense subset of A_0. As in the proof of Theorem 3, we can show that

$$R_\theta'(D) \le R_\theta(D) \qquad (66)$$

for all $\theta \in \Lambda$ where $R_\theta'(\cdot)$ is the rate-distortion function for source $(A,\mathcal{A},\mu_\theta)$ and reproducing alphabet A_0'. Under the hypothesis of Corollary 1, this implies $R_\theta'(D) < \infty$ for all $\theta \in \Lambda$. To apply Theorem 5 to the family of sources $\{(A,\mathcal{A},\mu_\theta)|\theta \in \Lambda\}$ with reproducing alphabet A_0', we must in addition show that there is a $b' \in A_0'$ for which

$$E_\theta[\rho(X_0,b')]^2 < \infty,\ \theta \in \Lambda. \qquad (67)$$

We show (67) in the case of Example 1. Since ρ is bounded and $\rho(x,\cdot)$ is continuous for each $x \in A_0$, ρ^2 is bounded and $(\rho(x,\cdot))^2$ is continuous for each $x \in A_0$. From Condition (b), there is a $b \in \hat{A}_0$ for which $E_\theta[\rho(X_0,b)]^2 < \infty$, and from Condition (a), there is a $b' \in A_0'$ for which $E_\theta[\rho(X_0,b')]^2 < E_\theta[\rho(X_0 b)]^2 + 1 < \infty$, establishing (67).

The proof of (67) in the case of Example 2 is similar in spirit.

We can now apply Theorem 5 to obtain for $\epsilon > 0$, a sequence of codes $\{\underline{B}_n\}$ and for each $\theta \in \Lambda$ an integer $N(\theta)$ such that for $n \geq N(\theta)$,

$$\bar{\rho}_n(\theta) \leq D + \epsilon$$

and

$$\bar{r}_n(\theta) \leq (\ln 2)^{-1}[R_\theta'(D) + \epsilon].$$

Applying (66) gives the result of the theorem.

Appendix B

Proof of Theorem 6. Let $D \geq 0$ and $\epsilon > 0$. Let ρ_m be a maximum value of distortion. First choose $a > 1$ large enough and then $\delta > 0$ small enough such that

$$a\delta + \rho_m \left((a-1)^{-1}2 + \delta\right) < \epsilon \tag{68}$$

$$(\ln J + \delta)\left((a-1)^{-1}2 + \delta\right) < \epsilon/4 \tag{69}$$

and

$$\delta < \epsilon/4\ln 2 \tag{70}$$

Since $\{\mu_\theta | \theta \in \Lambda\}$ is totally bounded under $\bar{\rho}$, Λ can be partitioned into a finite collection of sets Λ_k, $k = 1,\dots K$ with the property that $\theta, \theta^* \in \Lambda_k$ implies $\bar{\rho}(\theta, \theta^*) < \delta$. For $k = 1,\dots K$ and integers $\ell = 1,\dots \lceil \delta \ln J \rceil$, define sets $\Lambda_{k,\ell} \subset \Lambda_k$ by

$$\Lambda_{k,\ell} = \{\theta \in \Lambda_k | R_\theta(D) \in [(\ell-1)\delta, \ell\delta)\}.$$

We reindex the family of sets $\Lambda_{k,\ell}$, $1 \leq k \leq K$, $1 \leq \ell \leq \lceil \delta \ln J \rceil$ to obtain a family of sets Λ_m, $m = 1,\dots M$. For $m = 1,\dots M$, label a parameter in Λ_m by θ_m; it follows that for $\theta \in \Lambda_m$, we have

$$\bar{\rho}(\theta, \theta_m) < \delta \tag{71}$$

and

$$|R_\theta(D) - R_{\theta_m}(D)| < \delta \tag{72}$$

Theorem 1 implies that for the finite family of sources $\{\mu_{\theta_m} | m = 1,\dots M\}$ there is an integer n and block codes of length n $C_n(m)$, $m = 1,\dots M$, satisfying

$$\mu_{\theta_m}\{\rho(\underline{X}^n | C_n(m)) > D + \delta\} < \delta \tag{73}$$

and

$$n^{-1}\lceil \log_2 |C_n(m)| \rceil < (\ln 2)^{-1}[R_{\theta_m}(D) + \delta] \tag{74}$$

for m = 1,... M. We incidentally assume n is large enough to satisfy

$$n^{-1}\lceil \log_2 M \rceil < \epsilon/4 \ln 2 \qquad (75)$$

We now use methods of [12] to combine the M block codes into a binary variable rate code. It follows from (74) that each code book $C_n(m)$ has a binary representation $B_n(m)$ of $|C_n(m)|$ binary words of length less than $n(\ln 2)^{-1}[R_{\theta_m}(D) + \delta]$ bits. Denote the correspondence between codewords of $B_n(m)$ and $C_n(m)$ by the bijective mapping $V_{n,m}: B_n(m) \to C_n(m)$. Prefix each codeword in $B_n(m)$ with the $\lceil \log_2 M \rceil$ bit binary representation of the integer m, and denote the resulting codebook by $\hat{B}_n(m)$. The mapping $\hat{V}_{n,m}: \hat{B}_n(m) \to C_n(m)$ is the bijective mapping for which $\hat{V}^{-1}_{n,m}(\underline{c})$ is just the prefixed version of $V^{-1}_{n,m}(\underline{c})$ for $\underline{c} \in C_n(m)$.

Let $B_n = \bigcup_m \hat{B}_n(m)$ and define the mapping $V_n: B_n \to \hat{A}_0^n$ by $V_n(\underline{b}) = \hat{V}_{n,m}(\underline{b})$ if $\underline{b} \in \hat{B}_n(m)$. The mapping $U_n: A_0^n \to B_n$ is defined as follows:

1) If $\min_m \rho(\underline{x}^n | C_n(m)) \le D + a\delta$, let $U_n(\underline{x}^n)$ be a minimal length codeword $\underline{b} \in B_n$ which satisfies $\rho(\underline{x}^n, V_n(\underline{b})) \le D + a\delta$.

2) If $\min_m \rho(\underline{x}^n | C_n(m)) > D + a\delta$, let $U_n(\underline{x}^n)$ be any codeword $\underline{b} \in B_n$.

Then $\underline{B}_n = (B_n, U_n, V_n)$ forms a binary variable rate code on blocks of length n.

In the remainder of the proof, we fix $\theta \in \Lambda$. It follows that $\theta \in \Lambda_m$ for some integer m. It follows from the code construction that the average distortion satisfies

$$\bar{\rho}_n(\theta) \le D + a\delta + \rho_m \mu_\theta\{\rho(\underline{X}^n | C_n(m)) > D + a\delta\} \qquad (76)$$

and using (72), the average rate satisfies

$$\bar{r}_n(\theta) \le (\ln 2)^{-1}[R_\theta(D)+\delta] + (\ln 2)^{-1}(\ln J+\delta)\mu_\theta\{\rho(\underline{X}^n|C_n(m)) > D+a\delta\} + n^{-1}\lceil \log_2 M\rceil \tag{77}$$

To proceed further, we need the following lemma, whose proof we defer.

<u>Lemma 5</u>. $\mu_\theta\{\rho(\underline{X}^n|C_n(m)) > D + a\delta\} < (a-1)^{-1}\, 2 + \delta.$ (78)

Using (78) in (76) gives

$$\bar{\rho}_n(\theta) \le D + a\delta + \rho_m((a-1)^{-1}2 + \delta)$$

and by (68)

$$\bar{\rho}_n(\theta) < D + \epsilon. \tag{79}$$

Using (78) in (77) gives

$$\bar{r}_n(\theta) \le (\ln 2)^{-1}[R_\theta(D) + \delta] + (\ln 2)^{-1}(\ln J + \delta)((a-1)^{-1}2 + \delta) + n^{-1}\lceil \log_2 M\rceil$$

and by (69), (70), and (75)

$$\bar{r}_n(\theta) < (\ln 2)^{-1}[R_\theta(D) + \epsilon] \tag{80}$$

Since (79) and (80) hold for $\theta \in \Lambda$ and $n \ge n^*$, the theorem will be established with the following.

<u>Proof of Lemma 5</u>. From the theory of the $\bar{\rho}$ distance, there is a joint measure p^n on $A_0^n \times A_0^n$ having μ_θ^n and $\mu_{\theta_m}^n$ as marginals for which

$$\begin{aligned}\int_{A_0^n \times A_0^n} \rho(\underline{x}^n, \underline{y}^n)p^n(d\underline{x}^n, d\underline{y}^n) &\le \bar{\rho}_n(\theta,\theta_m) + \delta \\ &\le \bar{\rho}(\theta,\theta_m) + \delta \\ &< 2\,\delta.\end{aligned} \tag{81}$$

where the last inequality follows from (71).

Define the set $S \subset A_0^n$ and the set $T \subset A_0^n$ by

$$S = \{\underline{x}^n | \rho(\underline{x}^n | C_n(m)) > D + a\delta\}$$

$$T = \{\underline{x}^n | \rho(\underline{x}^n | C_n(m)) \leq D + \delta\}$$

For $(\underline{x}^n, \underline{y}^n) \in S \times T$, $\rho(\underline{x}^n, \underline{y}^n) > (a-1)\delta$. Thus from (81)

$$2\delta > \int_{S \times T} \rho(\underline{x}^n, \underline{y}^n) p^n(d\underline{x}^n, d\underline{y}^n)$$

$$> (a-1)\delta \; p^n\{S \times T\} \tag{82}$$

From (73) and (82),

$$\begin{aligned} \mu_\theta^n\{S\} &= p^n\{S \times A_0^n\} \\ &= p^n\{S \times T\} + p^n\{S \times T^c\} \\ &< (a-1)^{-1} 2 + p^n\{A_0^n \times T^c\} \\ &= (a-1)^{-1} 2 + \mu_{\theta_m}^n\{T^c\} \\ &< (a-1)^{-1} 2 + \delta. \end{aligned}$$

<u>Proof of Proposition 2</u>.

This proof is the variable rate analogue to [8, Theorem 5.1]. For $\epsilon > 0$, choose n large enough such that

$$r^b(n, D, \theta) \leq R_\theta(D) + \epsilon/2 \tag{83}$$

for all $\theta \in \Lambda$. The above bound can be achieved for all $\theta \in \Lambda$ by considering codes from the smaller class C of D-bounded distortion codes $\underline{B}$ for which $\ell[U(\underline{x}^n)] < |A_0^n|$, $\underline{x}^n \in A_0^n$, since for any code $\underline{B}' \notin C$, there is a code $\underline{B}'' \in C$ of the same average distortion and no greater average rate gotten by changing B' into a Huffman code. Since $|C| < \infty$, we may choose k sufficiently large that

$$(nk)^{-1} \lceil \log_2 |C| \rceil < \epsilon/2 \tag{84}$$

and index the codes in C as $\underline{B}^i$, $i=1,\ldots,|C|$. For $i=1,\ldots,|C|$, let w_i be the $\lceil \log_2 |C| \rceil$ bit binary representation of i, and form the cancatenated binary variable rate codes $\underline{B}^i_{nk}$ consisting of binary codebook $B^i_{nk} = w_i \times [B^i]^k$ and the natural encoding and decoding mappings $U_{nk} = (U^i, U^i, \ldots, U^i)$ and $V_{nk} = (V^i, V^i, \ldots, V^i)$. Combine the codes to form a code $\underline{B}_{nk}$ as follows:

(a) $B_{nk} = \bigcup_{1 \le i \le |C|} B^i_{nk}$,

(b) $U_{nk}(\underline{x}^{nk}) = U^i_{nk}(\underline{x}^{nk})$ if

$\ell[U^i_{nk}(\underline{x}^{nk})] = \min \{\ell[U^j_{nk}(\underline{x}^{nk})] \mid j=1,\ldots,|C|\}$

and

$\ell[U^i_{nk}(\underline{x}^{nk})] < \min \{\ell[U^j_{nk}(\underline{x}^{nk})] \mid j=1,\ldots,i-1\}$,

(c) $V_{nk}(w_i, b) = V^i_{nk}(w_i, b)$, $i=1,\ldots,|C|$.

It is obvious from the code construction that the resulting code is a D-bounded distortion code and therefore

$$\bar{d}(\underline{B}_{nk}, \theta) \le D \tag{85}$$

for all $\theta \in \Lambda$. Also, since

$$\ell[U_{nk}(\underline{x}^{nk})] = \min_i \ell[U^i_{nk}(\underline{x}^{nk})],$$

we have

$$
\begin{aligned}
\bar{r}(\underline{B}_{nk},\theta) &= (nk)^{-1}E_\theta\{\min_i \ell[U^i_{nk}(\underline{X}^{nk})]\} \\
&\le (nk)^{-1}\min_i E_\theta\ \ell[U^i_{nk}(\underline{X}^{nk})] \\
&= \min_i n^{-1}E_\theta\ \ell[U^i(\underline{X}^n)] + (nk)^{-1}\lceil \log_2|C|\rceil \\
&< r^b(n,D,\theta) + \epsilon/2
\end{aligned}
\tag{86}
$$

for all $\theta \in \Lambda$. Using (83) completes the proof.

<u>Proof of Theorem 7</u>. The transition matrix of a binary first-order Markov source has the form

$$
\underline{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \tag{87}
$$

where $0 \le p,\ q \le 1$. If p and q are not both 0, the corresponding vector of stationary probabilities is

$$
\underline{\Pi} = \left(\frac{q}{p+q},\ \frac{p}{p+q}\right) \tag{88}
$$

The source $(A,\mathcal{A},\mu_\theta{*})$ given by

$$
\underline{P}{*} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
$$

is nonergodic and $\underline{\Pi}$ can be any probability vector.

As in [11], we define the matrix metric $|\theta_1 - \theta_2|$ between sources $\mu_{\theta_1} = (\underline{\Pi}^{(1)}, \underline{P}^{(1)})$ and $\mu_{\theta_2} = (\underline{\Pi}^{(2)}, \underline{P}^{(2)})$ by

$$|\theta_1 - \theta_2| = \max\{|\Pi_i^{(1)} - \Pi_j^{(2)}|, |P_{ij}^{(1)} - P_{ij}^{(2)}|, 1 \le i, j \le 2\}.$$

Under the matrix metric, Λ is compact and for any n both $R_{\theta,n}(D)$ and $r^b(n,D,\theta)$ are continuous functions on Λ since both depend continuously on the n-th order distribution of μ_θ which is in turn a continuous function of θ. As in [11], we can then show $R_\theta(D)$ is continuous on Λ. Since $r^b(D,\theta^*) = R_{\theta^*}(D) = 0$ and since by Proposition 1, $r^b(D,\theta) = R_\theta(D)$ for all other $\theta \in \Lambda$, $r^b(D,\theta)$ is continuous on Λ. As in [11], we can use the subadditivity of the continuous function $r^b(n,D,\theta)$ to show that $r^b(n,D,\theta)$ converges *uniformly to the continuous* function $r^b(D,\theta)$ on Λ. By Proposition 2, this implies strongly universal codes exist.

<u>Proof of Proposition 3</u>. Let $\delta > 0$. Since $\{R_{N,\theta}(\cdot) | N \ge 1, \theta \in \Lambda\}$ is a family of convex strictly decreasing functions of the argument bounded above by lnJ and below by 0, there is an $\epsilon' > 0$ such that for any ϵ, $0 \le \epsilon \le \epsilon'$,

$$R_{N,\theta}(D) \le R_{N,\theta}(D + \epsilon) + \delta \tag{89}$$

for all $N \ge 1$, $\theta \in \Lambda$. By assumption, there is an integer n and a binary variable rate code on blocks of length n, $\underline{B}_n$, such that

$$\bar{\rho}_n(\theta) \le D + \epsilon \tag{90}$$

and

$$\bar{r}_n(\theta) \le (\ln 2)^{-1}[R_\theta(D) + \epsilon] \tag{91}$$

for all $\theta \in \Lambda$. The converse coding theorem applied to (90) gives

$$(\ln 2)^{-1} R_{n,\theta}(D + \epsilon) \le \bar{r}_n(\theta) \tag{92}$$

for all $\theta \in \Lambda$. Combining (89), (91), and (92) gives

$$R_{n,\theta}(D) \le R_\theta(D) + \epsilon + \delta$$

for all $\theta \in \Lambda$. Since δ and ϵ can be taken arbitrarily small, the theorem is established.

<u>Proof of Proposition 4</u>. Fix

$$D \in (0, \min\{J^{-1} \sum_{y \in \hat{A}_0} \rho(x,y) \mid x \in A_0\}).$$

For any positive integer n and any p, $0 < p < 1$, consider the n-th order Markov source with alphabet A_0 and transition probability

$$p(x_k \mid x_{k-1}, \ldots, x_{k-n}) = \begin{cases} p & \text{if } x_k = x_{k-n} \\ \frac{1-p}{J-1} & \text{otherwise} \end{cases}$$

It can be shown that the stationary distribution for this transition probability is

$$p(\underline{x}^n) = J^{-n} \tag{93}$$

for all $\underline{x}^n \in A_0^n$. Let $(A\mathcal{A}, \mu_{\theta(n,p)})$ be the stationary ergodic source corresponding to this transition probability and stationary distribution. We have

$$R_{\theta(n,p)}(0) = -p\ln p - J(1-p)\ln(1-p) \tag{94}$$

Let $(A,\mathcal{A},\mu_{\theta*})$ be the memoryless source which assigns probability mass J^{-1} to each $x \in A_0$. Since $R_{\theta*}(D) > 0$, using (94) we can choose $\tilde{p}$ small enough such that $R_{\theta(n,\tilde{p})}(0) < R_{\theta*}(D)/2$ implying

$$R_{\theta(n,\tilde{p})}(D) < R_{\theta*}(D)/2.$$

But from (93), $R_{n,\theta(n,\tilde{p})}(D) = R_{\theta*}(D)$ so that

$$R_{n,\theta(n,\tilde{p})}(D) - R_{\theta(n,\tilde{p})}(D) > R_{\theta*}(D)/2.$$

Since $\theta(n,\tilde{p}) \in \Lambda$ for all $n \geq 1$, we have shown that $\{R_{n,\theta}(D)\}$ does not converge uniformly on Λ which by Proposition 3 implies that Λ is not strongly universally encodable.

<u>Proof of Proposition 5</u>. We use proof by contradiction. Fix $\epsilon > 0$, $D \geq 0$, and assume a binary variable rate code on blocks of length n, $\underline{B}_n = (B,U,V)$, has been found for which

$$\bar{\rho}_n(\theta) \leq D + \epsilon \tag{95}$$

and

$$\bar{r}_n(\theta) \leq (\ln 2)^{-1}[R_\theta(D) + \epsilon] \tag{96}$$

for all $\theta \in \Lambda$. Select a countable set $\{a_1,a_2,\ldots\} \subset A_0$ for which $\rho(a_i,a_j) > 2(D + \epsilon)$ for $i \neq j$. For all positive integers k, let $(A,\mathcal{A},\mu_{\theta_k})$ be the source satisfying $\mu_{\theta_k}\{\ldots.a_k,a_k,a_k,\ldots\} = 1$, and let $\underline{a}_k^n$ be the n-tuple of a_k's.

<u>Claim</u> . $U(\underline{a}_i^n) \neq U(\underline{a}_j^n)$ for $i \neq j$.

<u>Proof of Claim</u>. If $U(\underline{a}_i^n) = U(\underline{a}_j^n)$, then

$$V[U(\underline{a}_i^n)] = V[U(\underline{a}_j^n)]. \tag{97}$$

But (95) implies $\rho(\underline{a}_i^n, V[U(\underline{a}_i^n)]) \le D + \epsilon$ and $\rho(\underline{a}_j^n, V[U(\underline{a}_j^n)]) \le D + \epsilon$, which using (97) gives $\rho(\underline{a}_i^n, \underline{a}_j^n) \le 2(D + \epsilon)$, a contradiction.

Since B is a **binary prefix condition codebook, we have** using the claim

$$1 \ge \sum_{\underline{b} \in B} 2^{-\ell(\underline{b})} \ge \sum_{k=1}^{\infty} 2^{-\ell[U(\underline{a}_k^n)]}$$

But from (96), $\bar{r}_n(\theta_k) = n^{-1} \ell[U(\underline{a}_k^n)] \le \epsilon$, which used in the above inequality gives

$$1 \ge \sum_{k=1}^{\infty} 2^{-n\epsilon},$$

a contradiction. This establishes the theorem.

References

1. C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal, pp. 379-423, 623-656, July and October, 1948.

2. C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," IRE National Convention Record, Part 4, pp. 142-163, 1959.

3. B. M. Fitingof, "Optimal coding in the case of unknown and changing message statistics," Problems of Information Transmission, vol. 2, no. 2, pp. 1-7, 1966.

4. J. Ziv, "Coding of sources with unknown statistics - Part I: Probability of encoding error," IEEE Transactions on Information Theory, vol. IT-18, pp. 384-389, May 1972.

5. L. D. Davisson, "Universal noiseless coding," IEEE Transactions on Information Theory, vol. IT-19, pp. 783-795, Nov. 1973.

6. R. M. Gray and L. D. Davisson, "Source coding without the ergodic assumption," IEEE Transactions on Information Theory, vol. IT-20, pp. 502-516, July 1974.

7. J. C. Oxtoby, "Ergodic sets," Bulletin of the American Mathematical Society, vol. 58, pp. 116-136, March 1952.

8. D. L. Neuhoff, R. M. Gray, and L. D. Davisson, "Fixed rate universal block source coding with a fidelity criterion," IEEE Transactions on Information Theory, vol. IT-21, pp. 511-523, Sept. 1975.

9. J. Ziv, "Coding of sources with unknown statistics--Part II: Distortion relative to a fidelity criterion," IEEE Transactions on Information Theory, vol. IT-18, pp. 389-394, May 1972.

10. R. M. Gray, D. L. Neuhoff, and P. C. Shields, "A generalization of Ornstein's $\bar{d}$ distance with applications to information theory," Annals of Probability, vol. 3, pp. 315-328, April 1975.

11. D. L. Neuhoff and P. C. Shields, "Universal codes for binary Markov sources," submitted to IEEE Transactions on Information Theory.

12. M. B. Pursley and L. D. Davisson, "Variable rate coding for nonergodic sources and classes of sources subject to a fidelity constraint," IEEE Transactions on Information Theory, vol. IT-22, pp. 324-337, May 1976.

13. M. B. Pursley and K. M. Mackenthun, "Variable rate coding for classes of sources with generalized alphabets," to appear IEEE Transactions on Information Theory, Sept. 1977.

14. J. C. Kieffer, "A generalization of the Pursley-Davisson-Mackenthun variable-rate coding theorem," submitted to IEEE Transactions on Information Theory.

15. L. D. Davisson and M. B. Pursley, "A direct proof of the coding theorem for discrete sources with memory," IEEE Transactions on Information Theory, vol. IT-21, pp. 310-317, May 1975.

16. T. Berger, Rate Distortion Theory--A Mathematical Basis for Data Compression. Englewood Cliffs, New Jersey: Prentice Hall, 1971.

17. D. J. Sakrison, "The rate of a class of random processes," IEEE Transactions on Information Theory, vol. IT-16, pp. 10-16, Jan. 1970.

18. I. A. Ibragimov, "Some limit theorems for stationary processes," Theory of Probability and its Applications, vol. VII, no. 4, pp. 349-382, 1962.

19. P. H. Diananda, "Some probability limit theorems with statistical applications," Proceedings Cambridge Philosophical Society, 49, pp. 239-246, 1953.

20. J. L. Doob, Stochastic Processes. New York: John Wiley, 1953.

21. V. A. Volkonskii and Yu A. Rozanov, "Some limit theorems for random functions, I," Theory of Probability and its Applications, vol. IV, no. 2, pp. 178-197, 1959.

22. A. N. Kolmogorov and Yu A. Rozanov, "On strong mixing conditions for stationary Gaussian processes," Theory of Probability and its Applications, vol. V, no. 2, pp. 204-208, 1960.

23. Yu A. Rozanov, Stationary Random Processes. San Francisco: Holden-Day, 1967.

24. J. L. Doob, "The elementary Gaussian process," Annals of Mathematical Statistics, vol. 15, pp. 229-282, 1944.

25. T. E. Harris, "On chains of infinite order," Pacific Journal of Mathematics, vol. 5, pp. 707-725, 1955.

26. D. L. Neuhoff, "Optimal source coding for stationary nonergodic sources," Thirteenth Annual Allerton Conference on Circuit and System Theory, Oct. 1-3, 1975.

Vita

Ken Mackenthun was born in Madison, Wisconsin on June 29, 1949. He attended Case Western Reserve University (B.A. 1972), George Washington University (B.S. 1974), and the University of Illinois (M.S. 1976). For the first year of graduate study at the University of Illinois, he received a University of Illinois Fellowship. His M.S. thesis was published as "Variable rate source coding," Coordinated Science Laboratory Report R-710, December 1975. He has co-authored the following articles with Dr. M. B. Pursley: "Variable rate coding for classes of sources with generalized alphabets," (to appear IEEE Transactions on Information Theory, Sept. 1977); "Recent results in variable rate source coding subject to a fidelity constraint,"; "Variable rate universal block source coding subject to a fidelity constraint," (submitted to IEEE Transactions on Information Theory, Oct. 1976); and "Variable-rate, strongly- and weakly-universal source coding."